Some Results on Algebraic and Geometric Characterization of Linear Systems Models for Time Series Analysis

João José de Farias Neto
Institute for Advanced Studies - IEAv-CTA
Divisão de Informática, joaojfn@ieav.cta.br
Rod. dos Tamoios, km 5,5, Cep - 12.228-840
São José dos Campos, SP, Brazil

February 1, 2008

Abstract

It is shown that in the multivariate case the orders p, of the AR part, and q, of the MA part, are not invariants of the time series. Thus, it is concluded that it only makes sense to define the class of ARMA(p,p)- irreducible models, where p is the biggest of the system's Kronecker indices. This class is shown not to be a differentiable manifold, but to contain one, which is a generic subset of systems with all Kronecker indices equal to p. A formula which gives the metric tensor for riemannian manifolds of linear systems as a line integral in the complex plane is introduced for deterministic and stochastic cases and some tensors are obtained with it.

MSC-class: 93C05; 93A30; 93B29; 93B30

1 Introduction

Since the publication of Box and Jenkins book [3], the statistical methodology proposed by these authors has been disseminated and widely used to build mathematical models of time series. These models are essentially discrete time differential equations (thus, indeed difference equations), of which the order and the parameters (supposed constant) are determined as a function only of a unique sample of its output sequence - the time series - which is considered as a particular realization of a gaussian stochastic process resulting from a forced term (the input of the system) which is also considered as an unobservable gaussian process. The orders p of the difference equation and q of the input noise are identified with the aid of correlation and partial correlation functions.

In the beginning of the 80's, Tiao and Tsay [16] , of the statistics department of Wisconsin University, proposed a methodology analogous to Box and Jenkin's to the multivariate case, that is, the one in which several time series are treated simultaneously, being viewed as a vector time series. The advantage of this approach is that the separate treatment of the series would not take in consideration their interrelations, which otherwise would allow the building of a model of greater predictive capacity. The proposed methodology is based upon a direct generalization of the correlation and partial correlation functions, to identify the orders p and q of the model.

It happens, nevertheless, that there is not a method to determine isolately the orders p and q of a multivariate ARMA(p,q) model nor is it possible to come up with one, simply because, contrarily to the univariate case, a multivariate ARMA model has not independent intrinsic p and q orders. This will be shown using the polynomial representation of linear systems and unimodular matrices.

The necessary correction is also proposed, with the introduction of the ARMA(p,p)-irreducible models. Given the formal resemblance between this new class of models and the class of linear systems of a common McMillan degree, a natural question is if this new class also constitutes a differentiable manifold. If this was the case, a whole new set of parametrizations would be unveiled, as it happened with the so-called overlapping parametrizations.

Since the space of all linear dynamical systems with a common McMillan degree and the generic set in the ARMA(p,p)-irreducible class are differentiable manifolds,

it makes sense to be able to do *identification on a manifold* (see references [12] and [15] for the basic results on this area). The resulting path in the systems space until convergence will be almost independent of the parametrization [5]. This could be termed *coordinate-free identification.*

To use this approach, the Riemannian metric tensor G must be computed, since the Riemannian gradient is given by $G^{-1}\nabla J$, that is, G^{-1} times the gradient of a convenient objective function J. Furthermore, the local geometric properties of the manifold are defined by G and, as a consequence, some of its global properties too, like for instance the geodesics equations, which are obtained by the integration of a set of partial differential equations based upon G.

In [12], pp.149-155, a recipe is given for the obtention of G. It consists on a formula for the norm of tangent vectors (which are systems derivatives) as a function of the matrices of a state space representation, needing the solution of a Lyapunov equation; this norm takes straightforwardly to the metric tensor. Identification algorithms which require the solution of Lyapunov or Riccati equations are common in the literature.

Here a formula for the metric tensor is obtained, which gives it directly as a functional of the system's transfer function. The generality of that formula enables the use of any representation besides state space (at least two other ones are known: ARMA and matrix pencil [2]) and doesn't require the solution of Lyapunov or Riccati equations. Particularized formulas for ARMA and state space representations are obtained, using overlapping parametrizations.

2 Definitions

An ARMA(p,q) (Auto-Regressive Moving-Averages) model for a vector time series $\{y_t\}$, $y \in \mathcal{R}^m$, t = 0,1,2,3..., is an equation of the type

$$A_0 y_t + A_1 y_{t-1} + ... + A_p y_{t-p} = B_0 \varepsilon_t + B_1 \varepsilon_{t-1} + ... + B_q \varepsilon_{t-q} \tag{1}$$

where the A_i and the B_i are m×m square matrices and $\{\varepsilon_t\}$ is a gaussian white noise of null mean and covariance R.

In the frequency domain (z-transform), the model becomes

$$\left(A_0 + A_1 z^{-1} + ... + A_p z^{-p}\right) Y(z) = \left(B_0 + B_1 z^{-1} + ... + B_p z^{-q}\right) \epsilon(z) \tag{2}$$

where

$$Y(z) = \sum_{i=0}^{\infty} y_i z^{-i} \tag{3}$$

and

$$\epsilon(z) = \sum_{i=0}^{\infty} \varepsilon_i z^{-i} \tag{4}$$

and z is a complex variable (if z is restricted to the unitary circle in the complex plane, it will have module 1 and, so, will be able to be written as $z = e^{\omega \mathrm{i}}$,with i=$\sqrt{-1}$ and $\omega \in \mathcal{R}$; in this case. the z-transform is reduced to the discrete Fourier transform, which justifies the expression "frequency domain ").

Now, let

$$A(z) = A_0 + A_1 z^{-1} + ... + A_p z^{-p} \tag{5}$$

$$B(z) = B_0 + B_1 z^{-1} + ... + B_q z^{-q} \tag{6}$$

Then, the frequency domain equation can be written as

$$Y(z) = A(z)^{-1} B(z) \epsilon(z) \tag{7}$$

(In the interesting cases, the nondegenerate ones, A(z) is invertible).
The so called transfer function of the model is

$$H(z) = A(z)^{-1} B(z) \tag{8}$$

Thus,

$$Y(z) = H(z) \epsilon(z) \tag{9}$$

It can be proved that, if $\{\varepsilon_t\}$is a gaussian stochastic process, $\{y_t\}$ will also be.

As the definition of H(z) implies that its elements will be fractions whose numerators and denominators are polynomials in z, that is, rational functions, the conclusion is that ARMA models represent linear dynamical systems, that is, systems whose input-output relation is of the type

$$y_t = \sum_{i=0}^{\infty} H_i \varepsilon_{t-i} \tag{10}$$

where $H(z) = \sum_{i=0}^{\infty} H_i z^{-i}$

being the H_i p X p square matrices known as the system's Markov parameters (or weighting sequence). A linear dynamical system is defined by its sequence $\{H_i\}$ of Markov parameters, so that each system s can be viewed as a point (or vector) in the Hilbert space of these sequences, defined by:

Sum of two systems: $s_3 = s_1 + s_2$ is the system such that $H_i^{(3)} = H_i^{(1)} + H_i^{(2)}$, i = 0,1,2,....

Product of a system by a scalar α : $s_2 = \alpha s_1$ is the system such that $H_i^{(2)} = \alpha H_i^{(1)}$, i = 0,1,2,...

Internal product between two systems:

$<s_1,s_2> = \mathrm{tr}\left[\sum_{i=0}^{\infty} H_i^{(1)} (H_i^{(2)})^T\right]$,

where T represents matrix transposition.

Norm of a system:$||s|| = \sqrt{< s_1, s_2 >}$

Distance between two systems:$d(s_1,s_2) = ||s_1 - s_2||$

These definitions in terms of $\{H_i\}$ are equivalent to the following ones in terms of H(z):

Sum of two systems: $s_3 = s_1 + s_2$ is the system such that $H_3(z) = H_1(z) + H_2(z)$
.

Product of a system by a scalar α : $s_2 = \alpha s_1$ is the system such that $H_2(z) = \alpha H_1(z)$, i = 0,1,2,...

Internal product between two systems:
$<s_1, s_2> = \frac{1}{2\pi i} \oint_C tr\left[H_1(z)H_2^T(z^{-1})\right] z^{-1}dz,$
where $i=\sqrt{-1}$and C is the unit circle in complex plane.

Norm of a system:$||s|| = \sqrt{< s_1, s_2 >}$

Distance between two systems:$d(s_1,s_2) = ||s_1 - s_2||$

The system will be stable if the minimum common multiple of the denominators of the elements of H(z) (expressed as a rational matrix with all polynomials in z, not in z^{-1}(the B operator of Box & Jenkins)) has all of its roots inside the unit circle in the complex plane. The system's stability is a condition that guarantees that to a stationary input process $\{\varepsilon_t\}$ there corresponds a stationary output process $\{y_t\}$, stationarity here understood as time invariance of all the moments of the stochastic process.

Let, now, the system's Hankel matrix be defined by

$$\mathcal{H} = \begin{bmatrix} H_1 & H_2 & H_3 & ... \\ H_2 & H_3 & H_4 & ... \\ ... & ... & ... & ... \end{bmatrix} \tag{11}$$

and search its lines top down, retaining only the ones which are linearly independent with the preceding ones. Associated to $\mathcal{H}$ there are m (the number of components of the output) positive integers, the so called Kronecker indices of the system; the index n_i is the number of retained lines corresponding to the i-th component of the vector time series $\{y_t\}$. The so called McMillan degree of the system is the rank of $\mathcal{H}$, which will be denoted by n. It is clear, then, that

$$n = \sum_{i=1}^{m} n_i \tag{12}$$

3 Why aren't p and q invariant

In the univariate case, m=1, A(z) e B(z) are polynomials in z^{-1}. Adopting the convention of using low case letters for scalars, the transfer function can be written as

$$h(z) = \frac{b(z)}{a(z)} \tag{13}$$

If b(z)=p(z)c(z) and a(z)=p(z)d(z), where p(z), c(z) and d(z) are polynomials in z^{-1}, the common factor p(z) can be cancelled, so that the same system (same transfer function) can be represented by the model

$$c(z)Y(z) = d(z)\epsilon(z) \tag{14}$$

If p was the degree of a(z), q of b(z) and r of p(z), the new model will be an ARMA(p-r,q-r). Thus, cancelling all the common factors of a(z) and b(z), an ARMA(p*,q*) with p* and q* minimal is obtained. It is obvious that any increase or reduction in p has, necessarily, to be accompanied by the same increase or reduction in q, since increases or reductions require adding or cancelling of common factors in the fraction h(z). The conclusion is that, in the scalar case, p* and q* are individually invariant, that is, given a system, it makes sense to refer to *the* order p* of its AR part and *the* order q* of its MA part and trying to identify them.

In the vector case, nevertheless, this is not true. In this case, the transfer function is

$$H(z) = A(z)^{-1}B(z) \tag{15}$$

If A(z)=P(z)C(z) and B(z)=P(z)D(z), with P,C e D matrix polynomials, then

$$H(z) = C(z)^{-1}P(z)^{-1}P(z)D(z) = C(z)^{-1}D(z) \tag{16}$$

Now, C=$\mathrm{P}^{-1}A$ and D=$\mathrm{P}^{-1}B$. If C and D are to be polynomial (so that CY=Dϵ be an ARMA model), it is necessary that all elements of A and B be divisible by detP (since $\mathrm{P}^{-1} = \frac{1}{\det P}Cof(P)^T$). Let d(z) be the maximum common divisor of elements of A e B. Define

$$P^{-1} = \frac{1}{d(z)}I \tag{17}$$

where I is the identity matrix. Then, P=d(z)I is polynomial and the only polynomial matrices M such that C=$\mathrm{M}^{-1}C$ and D=$\mathrm{M}^{-1}D$ are polynomial will be the ones whose determinant is not a polynomial in z^{-1} but a numerical constant; such polynomial matrices are known as unimodular matrices. Pairs (C,D) with such a property are said left coprime [13]. In the univariate case, after the cancelling of all common polynomial divisors, there remain only arbitrary numerical constants, which, canceled between numerator and denominator, don't change their degrees. In the multivariate case, there remain arbitrary unimodular matrices, which, once discounted, do alter the AR and MA degrees of the equation. That's why the following theorem can be stated:

THEOREM 3.1. *In the multivariate case, the minimal p and q are not individually invariant.*

Proof: Let s be a system representable by an ARMA(p,q+r) model of type [A(z),U(z)B(z)] irreducible (that is, with A and UB left coprime), with U unimodular, det$\mathrm{A}_p \neq 0$ e det$\mathrm{B}_q \neq 0$, where

$$U(z) = U_0 + U_1z^{-1} + ... + U_rz^{-r} \tag{18}$$

$$A(z) = A_0 + A_1z^{-1} + ... + A_pz^{-p} \tag{19}$$

(20) $$B(z) = B_0 + B_1 z^{-1} + ... + B_q z^{-q}$$

The restrictions over the determinants imply that the product of A(z) or B(z) by any polynomial matrix increase their degrees (for instance, U(z)B(z)=M_0 + $M_1 z^{-1} + ...M_{q+r} z^{-q+r}$, with $M_{q+r} = U_r B_q \neq O$, since $U_r B_q = O$ would imply on $U_r B_q B_q^{-1} = O B_q^{-1} => U_r = O$).

The coprimeness of (A,UB) implies on the impossibility of reducing the degrees of A and UB through the cancelling of a non-unimodular matrix P such that (A,UB)=(PM,PN). There remain only the unimodular ones.

Now, if U is unimodular, U^{-1} will be polynomial (and, by the way, unimodular too). So, $U^{-1}A$ will also be polynomial.

The conclusion is that $(U^{-1}A, U^{-1}UB) = (U^{-1}A, B)$ will be an ARMA(p+k,q) model, where k=degree(U^{-1}), irreducible, which will represent the same system. So, there are systems which have two irreducible ARMA representations with different degrees.

Q.E.D.

Consider, for instance, a system whose transfer function H(z) is itself a unimodular matrix. Then, this system has an ARMA(0,q) representation of the type

(21) $$Y(z) = H(z)\epsilon(z)$$

with

(22) $$H(z) = B_0 + B_1 z^{-1} + ... + B_r z^{-r}$$

But, in this case, $H(z)^{-1}$ is also polynomial, since

(23) $$H(z)^{-1} = \frac{1}{\det H(z)} Cof[H(z)]^T$$

since det(Hz) is a number and its cofactor matrix is polynomial.

Then, multiplying the ARMA(0,q) equation above by $H(z)^{-1}$, an ARMA(q,0) equation is obtained:

(24) $$H(z)^{-1}Y(z) = H(z)^{-1}H(z)\epsilon(z) = \epsilon(z)$$

Thus, there are two pairs, (0,q) and (q,0), of minimal orders (because nor p nor q can be smaller then zero) corresponding to the same system. Hannan and Deistler ([11], pg. 77), although don't call attention to this phenomenon, exhibit, *en passant*, an example with q=1.

Remark 1. *Although the systems built in the proof have a minimum p* and a minimum q*, this is not useful, since they don't have an ARMA(p*,q*) representation. In the case (0,q)-(q,0) exhibited, for instance, the system doesn't have an ARMA(0,0) representation (save for very special cases), which would represent a white noise. What would be useful for model building is the joint minimality of p and q, which would imply on the invariance of the structure of the minimum ARMA model.*

4 The ARMA(p,p)-irreducible class

Given the impossibility of representing any system by an ARMA(p,q) model with p and q jointly minimal, something that can be done is to represent it by an ARMA(p,p), that is, with p=q, such that p be the least possible integer. It can be proved (see [12], pg. 38) that this minimum value of p is equal to the largest of the Kronecker indices of the system.

The set of all systems representable by an ARMA(p,p) model with $A(z) = A_0 + A_1 z^{-1} + ... + A_p z^{-p}$, with p equal to their largest Kronecker index- that is, with p=max{n_i} - will be called *ARMA(p,p)-irreducible class* (the word irreducible here relates to the impossibility of reducing the value of p). It is a subset of the Hilbert space of linear systems. Some subsets of it are made up of systems with more than one ARMA(p,p)-irreducible model; so, the ARMA(p,p)-irreducible parametrization is not identifiable in the strict sense (although it is in Vajda's sense). (There are here, as always, two sets: the set of ARMA(p,p) models and the set of systems with max{n_i} $= p$, which are the image of those models in Hilbert's space. The parametrization which maps a set on the other one is said to be identifiable if it is biunivocal).

The problem posed here is if, analogously to the set of systems with a common $\sum_{i=1}^{m} n_i$, the set of systems with a common max{n_i} is a differentiable manifold; for, if that was the case, it would be possible to cover it with a set of charts, thus obtaining an overlapping parametrization which would be the most natural for the ARMA representation (in the state space representation, the natural thing is to treat with the set of systems with a common $\sum_{i=1}^{m} n_i$, which is the minimal dimension of the state space). To start the analysis, consider, firstly, an ARMA canonical form for systems with Kronecker indices n_i, i=1...m. It can be obtained by the following procedure:

Aline the components of the predictor y(t/t-1) (y(t) conditioned to the time series until time t-1), $y_1, y_2, ..., y_m$, from t-p to p. Next, adopt the procedure indicated for the example below, in which m=3, $n_1 = 2, n_2 = 2, n_3 = 2$:

$$\text{(25)} \qquad \left[\begin{array}{ccc} y_1 & y_2 & y_3 \\ * & * & * \\ & & \end{array}\right] \begin{array}{ccc} y_1 & y_2 & y_3 \\ * & o & * \\ * & * & * \end{array} \left[\begin{array}{ccc} y_1 & y_2 & y_3 \\ o & & o \\ * & & \end{array}\right]$$

where the leftmost matrix represents time t-2, the central one, t-1, and the right one, t. In the second line,the "x " indicate the components which enter in the selection and the "o " indicate the first time, in the search from left to right, that a component revealed to be linearly dependent with the precedent ones. The third line is the second one dislocated to the left, so as to have all L.D. components in the same column (the third one).

Now consider the third column. The first component (indicated in the second line) is L.D. with all components two time units behind and with the first and the third ones one time unit behind; so, in general, there will be coefficients corresponding to these components when writing down the model:

The second component depends of the first one in the same time and of all one

time unit behind, that is,

$y_2(t/t-1) = a_2 y_1(t/t-1) + b_2 y_1(t-1/t-2)$
$+c_2 y_2(t-1/t-2) + d_2 y_3(t-1/t-1)$

The third one depends on the first and the third ones one time unit behind and of all two time units behind, that is,

$y_3(t/t-1) = a_3 y_1(t-1/t-2) + b_3 y_3(t-1/t-2)$
$+c_3 y_1(t-2/t-3) + +d_3 y_2(t-2/t-3) + e_3 y_3(t-2/t-3)$

As to the MA part, all of its matrices are full, with the only restriction that, in each line of the equation, the degree of the MA part cannot be greater than the degree of the AR part.

Thus, the canonical structure becomes:

$$\begin{bmatrix} 1 & 0 & 0 \\ x & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} y_t + \begin{bmatrix} x & 0 & x \\ x & x & x \\ x & 0 & x \end{bmatrix} y_{t-1} + \begin{bmatrix} x & x & x \\ 0 & 0 & 0 \\ x & x & x \end{bmatrix} y_{t-2} =$$
$$= \varepsilon_t + \begin{bmatrix} x & x & x \\ x & x & x \\ x & x & x \end{bmatrix} \varepsilon_{t-1} + \begin{bmatrix} x & x & x \\ 0 & 0 & 0 \\ x & x & x \end{bmatrix} \varepsilon_{t-2}$$

Formally, what one has in the general case is (see [8], here modified to the stochastic case):

$$(26)\ y_i(t+n_i) = \sum_{j=1}^{m} \sum_{k=1}^{n_{ij}} a_{ijk} y_j(t+k-1) + \sum_{j=1}^{m} \sum_{k=1}^{n_i} b_{jik} \varepsilon_j(t+k-1) + \varepsilon_i(t+n_i)$$

where $n_{ij} = \begin{pmatrix} n_i, \text{ for i=j} \\ \min\{n_i+1, n_j\}, \text{ for i>j} \\ \min\{n_i, n_j\}, \text{ for i<j} \end{pmatrix}$

shifting, next, each equation in time, so that in the left member always appear $y_i(t)$.

The important to consider here are the following three facts:

1) $\det A_0 = 1$ always, so that one can multiply all the vector equation by A_0^{-1}, to obtain a monic model, that is, with $A_0 = I$ (mXm identity), without loss of generality.

2) Only when all Kronecker indices are equal is that all matrices - except for A_o, which will be the identity - are full. This is, thus, the case with the greatest number of free parameters: $2m^2p$; the dimension of its image M in the Hilbert space of systems will be, thus, also $2m^2p$. It is known that M is a differentiable manyfold, since each set of systems with the same m (number of components of y) and same set of Kronecker indices is the image of one of the maps of the chart which defines a differentiable manyfold of dimension 2mn [6], where $n = \sum_{i=1}^{m} n_i$ (when all indices are equal, one has $p = n_1 = ... = n_m$; thus n=mp and so $2m^2p = 2mn$). This is, consequently, the generic case of this parametrization.

3) In all the other cases, the augmented matrix $[A_p, B_p]$ will not be full rank (its product by an invertible matrix - A_0^{-1}, for instance - thus, also not).

Definition 1. *The ARMA(p,p)-irreducible parametrization will be defined by*

$$(27) \qquad y_t + A_1 y_{t-1} + ... + A_p y_{t-p} = \varepsilon_t + B_1 \varepsilon_{t-1} + ... + B_p \varepsilon_{t-p}$$

with all matrices, in principle, full. The word irreducible denotes the impossibility of reducing p. Thus, such parametrization includes all systems with max{n_i) $= p$ and only them.

From the properties of the canonical forms here exhibited, follows the conclusion that the image of the set of ARMA(p,p)-irreducible models in the space of systems is the union of a differentiable manyfold of dimension $2pm^2$ (corresponding to all systems with $n_1 = ... = n_m = p$) with sets of lower dimensions (corresponding to systems with some $n_i \neq p$). For easier references, it is convenient to state the following

THEOREM 4.1. *The generic sub-class of the ARMA(p,p)-irreducible parametrization is a differentiable manyfold of dimension $2pm^2$, where m is the number of components of* $\{y_t\}$.

Proof: The generic sub-class of this parametrization is the one in which all matrices are full and rank$[A_p, B_q] = m.$ But in this case the ARMA model is under the canonical form of systems with $n_1 = ... = n_m = p$, which, as is known [6], is a differentiable manifold of dimension $2pm^2$.

Q.E.D.

The problem here considered is of knowing if that union of systems sets, that is, the complete image, constitutes a differentiable manifold (in this case, its dimension would be $2pm^2$). This will occur if the points (systems) of the non-generic sets were regular under any coordinates system, which would require that the set of tangent vectors at each point spanned a space of dimension exactly equal to $2pm^2$. Unfortunately, the following theorem shows that this is not the case:

THEOREM 4.2. *The set of systems representable by the ARMA(p,p)-irreducible parametrization does not constitute a differentiable manyfold.*

Proof: Let s be a system representable by this parametrization with some $n_i \neq p$. Let $M(z) = I + Mz^{-1}$ with $M[A_p, B_p] = O$ (null matrix). For instance, in the case above illustrated$(m = 3, n_1 = 2, n_2 = 2, n_3 = 2)$,

$$M[A_p, B_p] = \begin{bmatrix} 0 & a & 0 \\ 0 & b & 0 \\ 0 & c & 0 \end{bmatrix} \begin{bmatrix} x & x & x & x & x & x \\ 0 & 0 & 0 & 0 & 0 & 0 \\ x & x & x & x & x & x \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{28}$$

and

$$M(z) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & a & 0 \\ 0 & b & 0 \\ 0 & c & 0 \end{bmatrix} z^{-1} = \begin{bmatrix} 1 & az^{-1} & 0 \\ 0 & 1 + bz^{-1} & 0 \\ 0 & cz^{-1} & 1 \end{bmatrix} \tag{29}$$

As in the systems with some $n_i \neq p$ the extended matrix $[A_p, B_p]$ is not full rank, there always is M with the properties above. In this case, the pair $[M(z)A(z), M(z)B(z)]$ will also represent an ARMA(p,p)-irreducible model with the same p, for the same system s, whatever be the values of a,b and c.

Let, now, $H_{\theta_i}(z)$, i=1...$2pm^2$, be the tangent vectors at a system representable by an ARMA(p,p)-irreducible parametrization, where the free parameters were denoted by $\{\theta_i, i = 1, 2, ...\}$. Since H(z) can be factored in

$H(z) = A(z)^{-1}B(z)$

being A(z) and B(z) the polynomial matrices

(30) $$A(z) = I + A_1 z^{-1} + ... + A_p z^{-p}$$

(31) $$B(z) = I + B_1 z^{-1} + ... + B_p z^{-p}$$

it follows that

(32) $$H_{\theta_i}(z) = A^{-1}(z) B_{\theta_i}(z) - A^{-1}(z) A_{\theta_i}(z) A^{-1}(z) B(z)$$

Or,

(33) $$H_{\theta_i}(z) = A^{-1}(z) \left(B_{\theta_i}(z) - A_{\theta_i}(z) H(z) \right)$$

Now, $B_{\theta_i}(z)$ and $A_{\theta_i}(z)$ are constant matrices (consider, for instance, the case m=2, p=2):

$y_t + \begin{bmatrix} \theta_1 & \theta_2 \\ \theta_3 & \theta_4 \end{bmatrix} y_{t-1} + \begin{bmatrix} \theta_5 & \theta_6 \\ \theta & \theta_8 \end{bmatrix} y_{t-2} =$

$= \varepsilon_t + \begin{bmatrix} \theta_9 & \theta_{10} \\ \theta_{11} & \theta_{12} \end{bmatrix} \varepsilon_{t-1} + \begin{bmatrix} \theta_{13} & \theta_{14} \\ \theta_{15} & \theta_{16} \end{bmatrix} \varepsilon_{t-2}$

Then, one has, for instance,

$A_{\theta_3}(z) = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} z^{-1}$

$B_{\theta_{13}}(z) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} z^{-2}$)

The matrix H(z) is unique and can be written in terms only of the free parameters of the canonical form of s.

However, if s is of the kind here considered (that is, such that some n$_i \neq p$), it will have an infinite number of representations [$\mathcal{A}$,$\mathcal{B}$], all of them being ARMA(p,p)-irreducible with the same p, as already claimed, of the type

$\mathcal{A}(z) = M(z)A(Z)$

$\mathcal{B}(z) = M(z)B(z)$

where A(z) and B(z) are the canonical matrices.

This doesn't change H(z), since

$\mathcal{H}(z) = \mathcal{A}(z)^{-1}\mathcal{B}(z) = [M(z)A(z)]^{-1}[M(z)B(z)]$

$= A(z)^{-1}M^{-1}(z)M(z)B(z) = A(z)^{-1}B(z) = H(z)$

But it does change A(z) in (33) and, so, H$_{\theta_i}$, i=1...2pm^2. In particular, it is always possible to define M(z) non-unimodular, maintaining the properties above (the case exhibited is an example: it suffices that b$\neq$ 0), such that M$^{-1}(z)$, when right-multiplied by A$^{-1}(z)$ in (33) introduce a free parameter (in this instance, b) in the denominator of H$_{\theta_i}$. Then, any change in the value of this parameter will result in a tangent vector (a system) which is L.I. with the remaining ones. Indeed, calling γ =($\gamma_1, \gamma_2, ..., \gamma_r$) the vector of free parameters of M(z) which appear in its determinant, one concludes that the space spanned by the tangent vectors at s is of infinite dimension (thus, different from 2pm^2), that is,

$\dim\{H_{\theta_i}(z)/\gamma \in R^r\} = \infty$

The conclusion is that points s of the kind here considered (that is, those ones not belonging to the generic sub-class) cannot be made regular under any coordinates system. Thus, the union of all classes of systems represented by the ARMA(p,p)-irreducible parametrization is not a differentiable manifold.

Q.E.D.

REMARK 2. *:The systems belonging to the non-generic case are in a position qualitatively similar to the vertex of a two-sided cone: it is not a regular point of the surface, but tangent vectors are perfectly definable on it, the anomaly being that they span of the R^3 instead of just a plane*

5 Riemannian Metric Tensor

Consider a parametrization (D,P) which defines a class of dynamical systems, so that the elements $h_{ij}(z)$ of the transfer function matrix H(z) which defines a system be expressed as a function of $\theta = (\theta_1, \theta_2, ..., \theta_d)$, that is, of the local coordinates of one of the maps of the differentiable manifold defined by (D,P), whose dimension will be denoted by "d ".

THEOREM 5.1. *The element (i,j) of the Riemannian metric tensor G of a dynamical linear system is given, in terms of its transfer function matrix H(z) and an adequate parametrization that expresses it as a function of a finite vector $\theta \in R^d$, by:*

$$g_{ij} = \frac{1}{2\pi \mathrm{i}} \oint_C tr\left[\frac{\partial H(z)}{\partial \theta_i}\frac{\partial H^T(z^{-1})}{\partial \theta_j}\right] z^{-1} dz$$

where i=$\sqrt{-1}$, C is the unit circle centered on the origin of the complex plane and tr stands for trace.

Proof:

Consider a system s belonging to the manifold defined by a

parametrization (D,P). Specifically, suppose that, in one of it's maps, s be represented by the vector of local coordinates $\theta = (\ \theta_1,\ \theta_2, ...,\ \theta_d)$. Let $\theta\,(v_1)$ and $\theta\,(v_2)$ be two curves of M passing by s. Let also s_1' and s_2' be the two derivatives of s along the two curves, respectively.

The Riemannian metric tensor G at s is, as already defined, the matrix such that

$${\theta_1'}^{\mathrm{T}} G\ \theta_2' = \mathrm{g}(s_1', s_2') = < s_1', s_2' > = \mathrm{tr}\left[\sum_{i=0}^{\infty} H_i^1 (H_i^2)^T\right]$$

where H_i^1 and H_i^2 , $i = 0, 1, 2, 3, ...,$ are the Markov parameters of the two tangent systems.

By Percival's formula (see [14], for the scalar case; the generalization used here is easily obtained), one has

$$\mathrm{tr}\left[\sum_{i=0}^{\infty} H_i^1 (H_i^2)^T\right] = \tfrac{1}{2\pi \mathrm{i}} \oint_C tr\left[H_1(z) H_2^T(z^{-1})\right] z^{-1} dz$$

where $\mathrm{H}_1(z)$ and $\mathrm{H}_2(z)$ are the transfer functions of the two tangent systems. Now, these transfer functions, because they refer to tangent systems along those curves (and recalling that stability of the systems implies convergence of the infinite series involved, thus justifying the commutation between summations and derivatives), are given by

$$H_k(z) = \sum_{i=0}^{\infty} \tfrac{dH_i}{dv_i} z^{-i} = \tfrac{d}{dv_i} \sum_{i=0}^{\infty} H_i z^{-i} = \tfrac{dH(z)}{dv_k} = \sum_{i=1}^{d} \tfrac{\partial H(z)}{\partial \theta_i} \tfrac{d\theta_i}{dv_k} =$$
$$= \sum_{i=1}^{d} \tfrac{\partial H(z)}{\partial \theta_i} \theta_i'^{(k)} \text{ , k=1,2}$$

where H(z) is the transfer function of s and $\frac{d\theta_i}{dv_k}$ was abbreviated to $\theta_i'^{(k)}$.
Substituting in the last integral, there results

$$\theta_1'^{\mathrm{T}} G\, \theta_2' = \tfrac{1}{2\pi \mathrm{i}} \oint_C tr\left[\left(\sum_{i=1}^{d} \tfrac{\partial H(z)}{\partial \theta_i}\theta_i'^{(1)}\right)\left(\sum_{i=1}^{d} \tfrac{\partial H(z^{-1})}{\partial \theta_i}\theta_i'^{(2)}\right)^T\right] z^{-1} dz$$

where $\theta_i'^{(1)} = [\theta_1']_i$, i=1,2,...,d
and $\theta_i'^{(2)} = [\theta_2']_i$, i=1,2,...,d

which specifies the components of vectors θ_1' and θ_2'.

The element g_{ij} of tensor G is obtained by
$\mathrm{g}_{ij} = e_i^T G e_j$
where $e_i = [0, ..., 0, 1, 0, ..., 0]$ with the "1 " in the i-th position.
Substitute, in the last integral, θ_1' by e_i and θ_2' by e_j.
The summations become:

$$\sum_{n=1}^{d} \tfrac{\partial H(z)}{\partial \theta_n}\theta_n'^{(1)} = \tfrac{\partial H(z)}{\partial \theta i} \qquad \text{and} \qquad \sum_{n=1}^{d} \tfrac{\partial H(z^{-1})}{\partial \theta_n}\theta_n'^{(2)} = \tfrac{\partial H(z^{-1})}{\partial \theta j}$$

and there results

$$\boxed{\mathrm{g}_{ij} = \tfrac{1}{2\pi \mathrm{i}} \oint_C tr\left[\tfrac{\partial H(z)}{\partial \theta_i}\tfrac{\partial H^T(z^{-1})}{\partial \theta_j}\right] z^{-1} dz}$$

Q.E.D.

Example 1. *For the scalar ARMA(1,1) parametrization,*
$y_t + a y_{t-1} = u_t + b u_{t-1}$, so that $\theta = (\theta_1, \theta_2) = (a, b)$, with $a, b \in (-1, 1)$ and $a \neq b$. The transfer function in this case is scalar, given by

$H(z) = \frac{z+b}{z+a} = \frac{z+\theta_1}{z+\theta_2}$

so that

$$\mathrm{g}_{ij} = \tfrac{1}{2\pi \mathrm{i}} \oint_C tr\left[\tfrac{\partial H(z)}{\partial \theta_i}\tfrac{\partial H^T(z^{-1})}{\partial \theta_j}\right] z^{-1} dz =$$
$$=. \tfrac{1}{2\pi \mathrm{i}} \oint_C \left(\tfrac{\partial}{\partial \theta_i}\tfrac{z+\theta_1}{z+\theta_2}\right)\left(\tfrac{\partial}{\partial \theta_j}\tfrac{z^{-1}+\theta_1}{z^{-1}+\theta_2}\right) z^{-1} dz$$

$i, j = 1, 2$

Thus,

$$\mathrm{g}_{11} = \frac{1}{2\pi \mathrm{i}} \oint_C \frac{z+b}{(z+a)^2} \frac{z^{-1}+b}{(z^{-1}+a)^2} z^{-1} dz = \frac{4ab-(b^2+1)(a^2+1)}{(a^2-1)^3}$$

$$\mathrm{g}_{12} = \mathrm{g}_{21} = \frac{1}{2\pi \mathrm{i}} \oint_C \frac{z+b}{(z+a)^2} \frac{1}{z^{-1}+a} z^{-1} dz = \frac{ab-1}{(1-a^2)^2}$$

$$\mathrm{g}_{22} = \frac{1}{2\pi \mathrm{i}} \oint_C \frac{1}{z+a} \frac{1}{z^{-1}+a} z^{-1} dz = \frac{1}{1-a^2}$$

As a result, the Riemannian metric tensor in this case is

$$\mathrm{G} = \begin{bmatrix} \frac{4ab-(b^2+1)(a^2+1)}{(a^2-1)^3} & \frac{ab-1}{(1-a^2)^2} \\ \frac{ab-1}{(1-a^2)^2} & \frac{1}{1-a^2} \end{bmatrix}$$

which agrees with the corresponding state-space case given in [1, pg. 222].

5.1 Overlapping Parametrizations

Consider the case in which $\mathrm{H}_0 = \mathrm{I}$ (identity matrix) and y_t and $\mathrm{u}_t \in R^m$, that is, r=m (number of inputs equal to number of outputs). There is no loss of generality, since all required is static redefinition of the inputs (or outputs) and the introduction of artificial dummy inputs (or outputs).

Consider, now, the differentiable manifold S_n of all systems of this kind with McMillan degree n fixed. For m$\neq$ 1, it is not possible to cover it with a unique map, so that the global chart is made up of a set of maps, each one of them characterized by a set of m natural numbers n_i, i=1,2,...,m, with $\sum_{i=1}^{m} n_i = n$. Call $\mathrm{M}_{n;\, n_1,n_2,\ldots,n_m}$ the corresponding map. Then, the coordinates of a system s described by this map are defined by the following procedure [10]:

Given the system's Hankel matrix

$$\mathcal{H} = \begin{bmatrix} H_1 & H_2 & H_3 & \ldots \\ H_2 & H_3 & H_4 & \ldots \\ \ldots & \ldots & \ldots & \ldots \end{bmatrix}$$

let H^i be its i-th block of m lines (for instance, $\mathrm{H}^2 = [H_2 H_3 H_4 \ldots])$ and let $h_{1i}, h_{2i}, \ldots, h_{mi}$ be the lines (of infinite size) of H^i. The fact that a system can be represented by the map $\mathrm{M}_{n;\, n_1,n_2,\ldots,n_m}$ means that the lines

$h_{11}, \ldots, h_{1n_1}; h_{21,\ldots,}h_{2n_2}; h_{m1}, \ldots h_{mn_m}$ constitute a base for the space of lines of $\mathcal{H}$. Thus, the lines $h_{1(n_1+1)}, \ldots, h_{m(n_m+1)}$ can be written as linear combinations of them:

$$h_{i(n_i+1)} = \sum_{j=1}^{m} \sum_{k=1}^{n_j} \alpha_{ijk} h_{jk} \tag{34}$$

i=1,...,m.

Now, call $h_{ij}(k)$ the k-th element of line h_{ij}. The 2mn numbers $\{\alpha_{ijk}, k = 1, ..., n_j; i, j = 1, ..., m,$ and $h_{ij}(k), i = 1, ..., m; j = 1, ...n_i; k = 1, ..., p\}$ are the system's coordinates according to map $M_{n;\, n_1, n_2, ..., n_m}$, that is, the components θ_i $(i = 1, 2, ..., 2nm)$ of the coordinate-vector θ which appear in the tensor's formula given by theorem 1. Let the first nm components of θ be the $\{\alpha_{ijk}\}$ and the last nm ones be the $\{h_{ij}(k)\}$.

Now, define the matrix K with n rows and m columns as in [10].
Each element of this matrix is one of the coordinates $h_{ij}(k)$.

5.2 Tensor for the ARMA representation

The autoregressive moving averages representation (ARMA)

$A_0 y_t + A_1 y_{t-1} + ... + A_p y_{t-p} = B_0 u_t + B_1 u_{t-1} + ... + B_p u_{t-p}$
where $y_t, u_t \in R^m$, A_i and B_i are $m \times m$ matrices and whose transfer function is $H(z) = A^{-1}(z)\, B(z)$
where $A(z) = A_0 z^p + A_1 z^{p-1} + ... + A_p$

and $B(z) = B_0 z^p + B_1 z^{p-1} + ... + B_p$
has an overlapping parametrization as defined in [10] in which
$p = \max\{n_i\}$
$a_{ii}(z) = z^{n_i} - \alpha_{iin_i} z^{n_i - 1} - ... - \alpha_{ii1}$
$a_{ij}(z) = -\alpha_{ijn_j} z^{n_j - 1} - \alpha_{ijn_j - 1} z^{n_j - 2} - \alpha_{ij1}, \quad i \neq j$

with the α_{ijk} already defined, and

B(z) = A(z) + M(z)K

where M(z) is a polynomial matrix whose entries are
pseudo-derivatives (in relation to z) of the entries of A(z) (if $f(z) = z^m + a_1 z^{m-1} + ... + a_m$, then $f^{(k)}(z) = z^{m-k} + a_1 z^{m-k-1} + ... + a_{m-k}$ is its pseudo-derivative of order k).

With the definition of vector θ given above, notice that matrices A(z) and M(z) are functions only of its first nm components, while matrix K is function only of its nm last ones. Thus,the following developments can be made:

$H(z) = A^{-1}(z)B(z)$

$\frac{\partial}{\partial\theta} H(z) = A^{-1}(z) \left[\frac{\partial}{\partial\theta} B(z) - \frac{\partial A(z)}{\partial\theta} H(z) \right]$

$\frac{\partial}{\partial\theta} B(z) = \frac{\partial}{\partial\theta} A(z) + \frac{\partial}{\partial\theta} M(z) K + M(z) \frac{\partial}{\partial\theta} K$

with

$\frac{\partial}{\partial\theta_i}A(z) = \frac{\partial}{\partial\theta_i}M(z) = 0$ for i > nm

$\frac{\partial}{\partial\theta_i}K = 0$ for i < nm+1

So, for i > nm, there results

$$\frac{\partial}{\partial\theta_i}H(z) = A^{-1}(z)\left[\frac{\partial}{\partial\theta_i}B(z) - \frac{\partial A(z)}{\partial\theta_i}H(z)\right] = A^{-1}(z)\frac{\partial}{\partial\theta_i}B(z) =$$
$$= A^{-1}(z)M(z)\frac{\partial}{\partial\theta_i}K$$

And, for i < nm+1,

$$\frac{\partial}{\partial\theta_i}H(z) = A^{-1}(z)\left[\frac{\partial}{\partial\theta_i}A(z) + \frac{\partial M(z)}{\partial\theta_i}K - \frac{\partial A(z)}{\partial\theta_i}H(z)\right] =$$

$$= A^{-1}(z)\left(\frac{\partial M(z)}{\partial\theta_i}K + \frac{\partial A(z)}{\partial\theta_i}\left[I - H(z)\right]\right)$$

The results above demand a separation in three cases, each one giving rise to a corresponding formula for the metric tensor.

Let I={1,2,...,nm} and J={nm+1,nm+2,...,2nm} be index sets.

Case 1: i,j∈J

$$\mathrm{g}_{ij} = \frac{1}{2\pi\mathrm{i}}\oint_C tr\left[A^{-1}(z)M(z)\frac{\partial K}{\partial\theta_i}\frac{\partial K^T}{\partial\theta_j}M^T(z^{-1})A^{-T}(z^{-1})\right]z^{-1}dz$$

Case 2: i,j∈I

$$\mathrm{g}_{ij} = \frac{1}{2\pi\mathrm{i}}\oint_C tr[A^{-1}(z)\left(\frac{\partial M(z)}{\partial\theta_i}K + \frac{\partial A(z)}{\partial\theta_i}\left[I - H(z)\right]\right).$$
$$.\left(K^T\frac{\partial M^T(z^{-1})}{\partial\theta_j} + \left[I - H^T(z^{-1})\right]\frac{\partial A^T(z^{-1})}{\partial\theta_j}\right)A^{-T}(z^{-1})]z^{-1}dz$$

Case 3: i∈I and j∈J

$$\mathrm{g}_{ij} = \frac{1}{2\pi\mathrm{i}}\oint_C tr$$
$$\left[A^{-1}(z)\left(\frac{\partial M(z)}{\partial\theta_i}K + \frac{\partial A(z)}{\partial\theta_i}\left[I - H(z)\right]\right)\frac{\partial K^T}{\partial\theta_j}M^T(z^{-1})A^{-T}(z^{-1})\right]z^{-1}dz$$

5.3 Tensor for the state-space representation

The state space representation

$x_{t+1} = Ax_t + Bu_t$
$y_t = Cx_t + u_t$

where $x_t \in R^n$ and $u_t, y_t \in R^m$
has the overlapping parametrization defined in [10], where
C is a matrix of zeros and ones, A is a sparse matrix in which the only non-nul variable entries are the α_{ijk} already defined and $B = K$ (as defined above).
Thus,the following calculations will provide explicit formulas for the tensor:

$H(z) = C(zI - A)^{-1}B + I$

and $\frac{\partial}{\partial \theta}$ H(z)=$\frac{\partial}{\partial \theta}$C(zI-A)$^{-1}B$

and, for i > nm,

$\frac{\partial}{\partial \theta_i}$ H(z)=$\frac{\partial}{\partial \theta_i}$C(zI-A)$^{-1}B$ =C(zI-A)$^{-1}\frac{\partial}{\partial \theta_i}B$

whilst, for i < nm+1,
$\frac{\partial}{\partial \theta_i}$ H(z)=$\frac{\partial}{\partial \theta_i}$C(zI-A)$^{-1}B$ =C$\left[\frac{\partial}{\partial \theta_i}(zI - A)^{-1}\right] B =$

$= C(zI - A)^{-1}\frac{\partial A}{\partial \theta_i}(zI - A)^{-1}$

Again, this generates three formulas for the tensor:

Case 1: i,j∈J (see definitions of I and J in the preceding section)

$$g_{ij} = \frac{1}{2\pi \mathrm{i}} \oint_C tr \left[C(zI - A)^{-1}\frac{\partial B}{\partial \theta_i}\frac{\partial B^T}{\partial \theta j}(z^{-1}I - A)^{-T}C^T\right] z^{-1}dz$$

Case 2: i,j∈I

$$g_{ij} = \frac{1}{2\pi \mathrm{i}} \oint_C tr \left[C(zI - A)^{-1}\frac{\partial A}{\partial \theta_i}(zI - A)^{-1}(z^{-1}I - A)^{-T}\frac{\partial A^T}{\partial \theta_j}(z^{-1}I - A)^{-T}C^T\right] z^{-1}dz$$

Case 3: i∈I and j∈J

$$g_{ij} = \\ = \frac{1}{2\pi \mathrm{i}} \oint_C tr \left[C(zI - A)^{-1}\frac{\partial A}{\partial \theta_i}(zI - A)^{-1}\frac{\partial B^T}{\partial \theta j}(z^{-1}I - A)^{-T}C^T\right] z^{-1}dz$$

Remark 3. *In the deterministic case, letting* $\theta_i = [K]_{kl}\, , i = nm{+}1, ..., 2nm,$ *with*

k = integer(i/m)-n+1 and l = i mod(m), which amounts to naming θ_i in lexicographic order in matrix K, one has $\frac{\partial K}{\partial \theta_i}\frac{\partial K^T}{\partial \theta_j} = \mathbb{O}$ (null matrix) whenever i mod(m) $\neq$ j mod(m) (x mod(y) meaning the rest of division of x by y). Thus, the display of the metric tensor formulas for the ARMA and state space representations shows that $g_{ij} = 0$ for certain pairs (i,j), since in case 1 the expression $\frac{\partial K}{\partial \theta_i}\frac{\partial K^T}{\partial \theta_j}$ ($=\frac{\partial B}{\partial \theta_i}\frac{\partial B^T}{\partial \theta j}$, as B=K) appears in both representations.

5.4 Stochastic Metric Tensor

Stochastic linear dynamical systems are here defined as those ones which, besides having an input channel for known vector sequences, are permanently being excited by a vector white noise, of which only the first and second moments are known, the sequence itself being unknown. Hanzon proposes an internal product for the tangent space of the manifold of these systems without known input, of fixed McMillan degree and equal number of components for the vectors of white noise input and colored noise output. The corresponding tensor formula is obtained here, based upon theorem 5.1.1. A simpler formula is also presented, derived from another suggested internal product by that author. Detailed calculations for their applications to some simple examples are shown.

A stationary time series will be considered here as a realization of a gaussian and stationary stochastic process $\{y_t\}$. It will be supposed that y has zero mean, y $\in \mathcal{R}^m$ and t =1,2,3,... , that is, time will be discrete. Thus, $\{y_t\}$ is totally characterized by its auto-covariance function

(35) $$\Gamma_i = E(y_t y_{t+i}^T), i = 0, 1, 2, ...,$$

where E(.) is the expectance operator over y 's probability distribution function and the superscript T stands for transposition. This stems from the fact that the joint probability density of a string of size I+1 of the process, that is, of the random vector $w^T = (y_t^T, y_{t+1}^T, ..., y_{t+I}^T)$ is given by ([9], pg. 90):

(36) $$f(w) = [(2\pi)^{m(I+1)} \det G]^{-1/2} e^{-\frac{1}{2}w^T G w}$$

where

(37) $$G = \begin{bmatrix} \Gamma_0 & \Gamma_1 & \Gamma_2 & ... & \Gamma_I \\ \Gamma_1^T & \Gamma_0 & \Gamma_1 & ... & \Gamma_{I-1} \\ ... & ... & ... & ... & ... \\ \Gamma_I^T & \Gamma_{I-1}^T & \Gamma_{I-2}^T & ... & \Gamma_0 \end{bmatrix}$$

Being $\{y\}$ stationary, it follows that $\Gamma_i = \Gamma_{-i}^T$.

The density above presupposes that G be nonsingular, which is equivalent to w's components being linearly independent.

There is a biunivocal relation between the autocovariance function of a stochastic process and its spectral density [7]. In this case, calling T(z) the spectral density, one has ([1], pg. 69):

$$(38) \qquad T(z) = \sum_{i=-\infty}^{\infty} \Gamma_i z^{-i}$$

Now, T(z) can always be factored as T(z)=H(z)RH^T(z^{-1}), where H(z) is a rational and stable matrix, with stable inverse, and R is symmetric positive definite.

Imposing, further, that H(z) be causal and H_0=I, there is a unique pair [H(z),R] correspondent to rational T(z) ([9], pg. 72).

H(z) may be interpreted as the transfer function of a linear system. As a result, {y} will be interpreted as the output of a stable linear system whose input is a unobserved white noise.

Hanzon [12] proposes an internal product for the manifold of stochastic systems with a common McMillan degree analogous to the deterministic case:

$$(39) \qquad < s_1, s_2 >= tr \left[\sum_{i=0}^{\infty} \Gamma_i^1 (\Gamma_i^2)^T \right]$$

where s_1 e s_2 are systems belonging to the tangent bundle of the manifold.

This is a metric of the covariance system, whose Markov parameters are $\{\Gamma_i\}$. So the immediate extension of theorem 5.1.1 to the stochastic case is:

Theorem 5.2. *A stochastic Riemannian metric tensor can be obtained by the formula*

$$(40) \qquad g_{ij} = \frac{1}{2\pi \mathrm{i}} \oint_C tr \left[\frac{\partial U(z)}{\partial \theta_i} \frac{\partial U^T(z^{-1})}{\partial \theta_j} \right] z^{-1} dz$$

where:

$$(41) \qquad U(z) = \sum_{i=0}^{\infty} \Gamma_i z^{-i},$$

Γ_i= E($y_t y_{t+i}^T$), $i = 0, 1, 2, ...,$ are the covariances of the stochastic process $\{y_t\}$generated by the stochastic system $[\{H_i\}, R]$,

tr stands for trace,

the superscript T indicates matrix transposition,

"i " is $\sqrt{-1}$ and

C is the unitary circle centered in the origin of the complex plane.

A more convenient metric, which can be expressed directly in terms of H(z) and R, is the one induced by the internal product defined over the two-sided infinite sequence $\{\Gamma_1, i = ..., -2, -1, 0, 1, 2, ...\}$ (reminding that $\Gamma_i = \Gamma_{-i}^T, \forall i \in \mathbb{Z}$):

$$(42) \qquad < s_1, s_2 >= tr \left[\sum_{i=-\infty}^{\infty} \Gamma_i^1 (\Gamma_i^2)^T \right]$$

Hanzon ([12], pg. 208) refers to this choice as *an also quite attractive possibility.* Defining T(z) = $\sum_{i=-\infty}^{\infty} \Gamma_i z^{-i}$, it follows the

Theorem 5.3. *A stochastic Riemannian metric tensor can be obtained by the formula*

$$g_{ij} = \frac{1}{2\pi \mathrm{i}} \oint_C tr \left[\frac{\partial T(z)}{\partial \theta_i} \frac{\partial T^T(z^{-1})}{\partial \theta_j} \right] z^{-1} dz \tag{43}$$

where T(z) = $\sum_{i=-\infty}^{\infty} \Gamma_i z^{-i}$ = H(z)RH^T(z^{-1}) is the spectral density of the output $\{y_t\}$ of the linear dynamic system whose transfer function is H(z) and whose input is a white noise $\{\varepsilon_t\}$ of covariance matrix R and zero mean.

Proof: Percival's formula

$$tr \left[\sum_{i=0}^{\infty} A_i^1 (A_i^2)^T \right] = \frac{1}{2\pi \mathrm{i}} \oint_C tr \left[A_1(z) A_2^T(z^{-1}) \right] z^{-1} dz \tag{44}$$

which was the kernel of the proof of theorem 5.1.1, relates a summation of the sequence $\{A_i, i \in \mathbb{N}\}$ to an integral of the z-transform A(z) = $\sum_{i=0}^{\infty} A_i z^{-1}$. It can also be stated as

$$tr \left[\sum_{i=-\infty}^{\infty} A_i^1 (A_i^2)^T \right] = \frac{1}{2\pi \mathrm{i}} \oint_C tr \left[A_1(z) A_2^T(z^{-1}) \right] z^{-1} dz \tag{45}$$

where A(z) is now defined by A(z) = $\sum_{i=-\infty}^{\infty} A_i z^{-1}$, relating, thus, the summation of the sequence $\{A_i, i \in \mathbb{Z}\}$ to an integral of this last z-transform.

The remainder of the proof is equal to that of theorem 5.1.

Q.E.D.

The relation between T(z) and U(z) is the following:

$$T(z) = U(z) + U^T(z^{-1}) - \Gamma_0 \tag{46}$$

5.5 Some examples of tensors for the stochastic case

Consider, as examples, the following stochastic systems (all of them scalar, that is, $y_t, x_t, \varepsilon_t \in \mathbb{R}$, and, to simplify, R=1):

5.5.1 In the state space representation

$$x_{t+1} = A x_t + B \varepsilon_t \tag{47}$$

$$y_t = C x_t + D \varepsilon_t \tag{48}$$

$y_t, x_t, \varepsilon_t \in \mathbb{R}$.

In this case, the covariance R of ε_t is a scalar and has a purely multiplicative effect on the tensor.

Example 1

To simplify, let R=1, D=0 and C=1. Then, the system reduces to

$$x_{t+1} = a x_t + b\varepsilon_t \tag{49}$$

$$y_t = x_t \tag{50}$$

whose transfer function is

$$h(z) = \frac{b}{z-a} \tag{51}$$

So, its spectral density is

$$T(z) = H(z)RH^T(z^{-1}) = \frac{b}{z-a}\frac{b}{z^{-1}-a} = -b^2\frac{z}{(-a+z)(-1+az)} \tag{52}$$

The partial derivatives become

$\frac{\partial T(z)}{\partial a} = b^2 z \frac{1-2az+z^2}{(-a+z)^2(-1+az)^2}$
$\frac{\partial T(z)}{\partial b} = -2b\frac{z}{(z-a)(-1+az)}$

Then, noticing that T(z)=T(z^{-1}),
$g_{11} = \frac{1}{2\pi \mathrm{i}} \oint_C tr\left[\frac{\partial T(z)}{\partial a}\frac{\partial T^T(z^{-1})}{\partial a}\right] z^{-1}dz =$
$= \frac{1}{2\pi \mathrm{i}} \oint_C \left[b^2 z\frac{1-2az+z^2}{(-a+z)^2(-1+az)^2}\right]^2 z^{-1}dz$
$g_{22} = \frac{1}{2\pi \mathrm{i}} \oint_C tr\left[\frac{\partial T(z)}{\partial b}\frac{\partial T^T(z^{-1})}{\partial b}\right] z^{-1}dz =$
$= \frac{1}{2\pi \mathrm{i}} \oint_C \left[-2b\frac{z}{(z-a)(-1+az)}\right]^2 z^{-1}dz$
$g_{12} = g_{21} = \frac{1}{2\pi \mathrm{i}} \oint_C tr\left[\frac{\partial T(z)}{\partial a}\frac{\partial T^T(z^{-1})}{\partial b}\right] z^{-1}dz$
$g_{12} = g_{21} =$
$= \frac{1}{2\pi \mathrm{i}} \oint_C \left[-2b\frac{z}{(z-a)(-1+az)}\right]\left[b^2 z\frac{1-2az+z^2}{(z-a)^2(-1+az)^2}\right] z^{-1}dz$

Calculating the integrals, the following tensor is obtained:

$$G = \begin{bmatrix} -2\frac{(2a^4+7a^2+1)b^4}{(a^2-1)^5} & 4\frac{ab^3(a^2+2)}{(a^2-1)^4} \\ 4\frac{ab^3(a^2+2)}{(a^2-1)^4} & -4\frac{b^2(a^2+1)}{(a^2-1)^3} \end{bmatrix} \tag{53}$$

This is an interesting example, because it is also exhibited in [12], pg. 223, using the metric of theorem 5.2 (but through another method, involving Ricati and Lyapunov equations and not a complex integral. The tensor obtained in that work is:

$$G = \begin{bmatrix} -\dfrac{(9a^2+1)b^4}{(a^2-1)^5} & 6\dfrac{ab^3}{(a^2-1)^4} \\ 6\dfrac{ab^3}{(a^2-1)^4} & -4\dfrac{b^2}{(a^2-1)^3} \end{bmatrix} \tag{54}$$

Notice that the denominators coincide with those of the tensor(53) , but not the numerators. The tensor (54) can be found by theorem 5.2 by the following process:

Expand the transfer function in power series

$$h(z) = \frac{b}{z-a} = bz^{-1}(1 + az^{-1} + a^2z^{-2} + ...) = bz^{-1} + abz^{-2} + a^2bz^{-3} + ... \tag{55}$$

to obtain the system's Markov parameters

$$h_0 = 0, h_1 = b, h_2 = ab, h_3 = a^2b, ..., h_i = a^{i-1}b \tag{56}$$

The sequence of covariances $\{\Gamma_k\}$ can be found by:

$$\Gamma_k = E(y_t y_{t+k}^T) = E\left[\sum_{i=0}^{\infty} H_i \varepsilon_{t-i} \left(\sum_{j=0}^{\infty} H_j \varepsilon_{t+k-j}\right)^T\right] =$$
$$= E\left[\sum_{i=0}^{\infty} H_i \varepsilon_{t-i} \sum_{j=0}^{\infty} \varepsilon_{t+k-j}^T H_j^T\right]$$
$$\Gamma_k = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} H_i E\left[\varepsilon_{t-i}\varepsilon_{t+k-j}^T\right] H_j^T = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} H_i R \delta_{j-k,i} H_j^T =$$
$$= \sum_{i=0}^{\infty} H_i R H_{i+k}^T$$

where $\delta_{j-k,i}$ is Kronecker's delta and reminding that $\mathrm{y}_t = \sum_{i=0}^{\infty} H_i \varepsilon_{t-i}$.

Substituting the Markov sequence of the example, comes

$$\Gamma_k = \sum_{i=1}^{\infty} a^{i-1} b^2 a^{i+k-1} \tag{57}$$

Then,

$$U(z) = \sum_{k=0}^{\infty}\left(\sum_{i=1}^{\infty} a^{i-1} b^2 a^{i+k-1}\right) z^{-k} = -b^2 \frac{z}{(a^2-1)(-a+z)} \tag{58}$$

Substituting in formula (40) of theorem 5.2, tensor (54) is obtained.

Notice that equation (46) is satisfied:

$U(z^{-1}) = -\frac{b^2}{z(a^2-1)\left(-a+\frac{1}{z}\right)}$

$$\Gamma_0 = \sum_{i=0}^{\infty} H_i R H_i^T = \sum_{i=1}^{\infty} a^{i-1} b^2 a^{i-1} = -\frac{b^2}{a^2-1} \tag{59}$$

$U(z) + U^T(z^{-1}) - \Gamma_0 =$
$= -b^2 \frac{z}{(a^2-1)(-a+z)} + \left(-\frac{b^2}{z(a^2-1)\left(-a+\frac{1}{z}\right)}\right) - \left(-\frac{b^2}{a^2-1}\right) =$
$= -z \frac{b^2}{(az-1)(-a+z)} = T(z)$ (see expression (52).

Example 2

The next example is of the important innovations model:

$$x_{t+1} = a x_t + b\varepsilon_t \tag{60}$$

$$y_t = x_t + \varepsilon_t \tag{61}$$

The transfer function in this case is

$$h(z) = \frac{z+b-a}{z-a} \tag{62}$$

The spectral density becomes

$$T(z) = h(z)h(z^{-1}) = (z+b-a)\,\frac{-1-bz+az}{(z-a)\,(-1+az)} \tag{63}$$

The result found by theorem 3 is:

$$g_{11} = -2\frac{(-a^4+7b^2a^2+1+a^6-a^2-4ba^5+2b^2a^4+4ba+b^2)b^2}{(a^2-1)^5}$$

$$g_{12} = g_{21} = 2\frac{b(a-2a^3+a^5-4ba^4+2b^2a^3+3ba^2+4b^2a+b)}{(a^2-1)^4} \tag{64}$$

$$g_{22} = -2\frac{a^4-2a^2+1-4ba^3+2b^2a^2+4ba+2b^2}{(a^2-1)^3}$$

To use theorem 5.2 semi-infinite metric, compute:
$h(z) = \frac{z+b-a}{z-a} = \frac{z-a}{z-a} + \frac{b}{z-a} = 1 + \frac{b}{z-a} = 1 + bz^{-1} + abz^{-2} + a^2bz^{-3} + ...$
Thus,

$$h_i = a^{i-1}b, i = 1, 2, 3, ... \tag{65}$$

$$h_0 = 1 \tag{66}$$

The general expressions

$\Gamma_k = \sum_{i=0}^{\infty} H_i R H_{i+k}^T$
$U(z) = \sum_{k=0}^{\infty} \Gamma_k z^{-k}$

become, for this case:

For $k = 1, 2, 3, ..$
$\Gamma_k = \sum_{i=0}^{\infty} H_i R H_{i+k}^T = \sum_{i=1}^{\infty} H_i R H_{i+k}^T + H_0 R H_k^T$

(67)
$$\Gamma_k = \sum_{i=1}^{\infty} a^{i-1}(b)^2 a^{i+k-1} + a^{k-1}b = -b\frac{a^k b - a^{1+k} + a^{k-1}}{a^2-1}$$

(68)
$$\Gamma_0 = \sum_{i=1}^{\infty} H_i R H_i^T + H_0 R H_0^T = \left(\sum_{i=1}^{\infty} a^{i-1}(b)^2 a^{i-1}\right) + 1 = \frac{-b^2 + a^2 - 1}{a^2 - 1}$$

Then, $U(z) = \sum_{k=0}^{\infty} \Gamma_k z^{-k} = \sum_{k=1}^{\infty} \Gamma_k z^{-k} + \Gamma_0$
$U(z) = \sum_{k=1}^{\infty} -b\frac{a^k b - a^{1+k} + a^{k-1}}{a^2-1} z^{-k} + \frac{-b^2+a^2-1}{a^2-1} = \frac{ba^2-b-b^2z-a^3+a^2z+a-z}{(a^2-1)(-a+z)}$

Having U(z), it is enough to apply theorem 5.2 formula (40), to get to the corresponding tensor:

(69)
$$g_{11} = -\left(a^6 - 4a^5 b - a^4 + 9b^2a^2 - a^2 + 4ba + 1 + b^2\right)\frac{b^2}{(a-1)^5(a+1)^5}$$
$$g_{12} = \left(a^5 - 4ba^4 - 2a^3 + 3ba^2 + a + 6ab^2 + b\right)\frac{b}{(a-1)^4(a+1)^4}$$
$$g_{22} = -\frac{-2a^2 + 4b^2 + 1 + a^4 + 4ba - 4a^3 b}{(a-1)^3(a+1)^3}$$

5.5.2 In the ARMA representation

The ARMA equations is:

(70)
$$A_0 y_t + A_1 y_{t-1} + ... + A_p y_{t-p} = B_0 u_t + B_1 u_{t-1} + ... + B_p u_{t-p}$$

The univariate ARMA(1,1) case will be exhibited here. It is defined by

Example 3

(71)
$$y_t + a y_{t-1} = \varepsilon_t + b\varepsilon_{t-1}$$

whose transfer function is

(72)
$$h(z) = \frac{z+b}{z+a}$$

The spectral density becomes

$$T(z) = h(z)h(z^{-1}) = \frac{z+b}{z+a}\frac{\frac{1}{z}+b}{\frac{1}{z}+a} = (z+b)\,\frac{1+bz}{(az+1)\,(a+z)} \tag{73}$$

The following tensor is obtained by theorem 3:

$$g_{11} = 2\tfrac{-43b^2a^2-5b^2+b^2a^6-13b^2a^4-2a^4-2b^4a^4+24b^3a^3-7a^2-7b^4a^2+16ba+16b^3a-1-b^4}{(a^2-1)^5}$$

$$g_{12} = -2\frac{-11ba-8ba^3+ba^5-2b^3a^3+15b^2a^2+5a^2-a^2b^3+1+3b^2}{(a^2-1)^4} \tag{74}$$

$$g_{22} = 2\frac{a^4-4a^2-1-2b^2a^2+8ba-2b^2}{(a^2-1)^3}$$

The computations to obtain the tensor prescribed by theorem 5.2 are:

$h(z) = \frac{z+b}{z+a} = \frac{z}{z+a} + \frac{b}{z+a}$

$\frac{z}{z+a} = \frac{1}{1+az^{-1}} = 1 - az^{-1} + a^2z^{-2} - a^3z^{-3} + ...$

$\frac{b}{z+a} = bz^{-1}\frac{1}{1+az^{-1}} = bz^{-1} - baz^{-2} + ba^3z^{-3} + ...$

Thus,

$$h(z) = 1 - (a-b)z^{-1} + a(a-b)z^{-2} - a^2(a-b)z^{-3} + ... \tag{75}$$

So,

$$h_i = a^{i-1}(a-b)(-1)^i, i = 1, 2, 3, ... \tag{76}$$

$$h_0 = 1 \tag{77}$$

Now, for $k = 1, 2, 3, ..$

$$\Gamma_k = \sum_{i=1}^{\infty} a^{i-1}(a-b)(-1)^i a^{i+k-1}(a-b)(-1)^{i+k} + a^{k-1}(a-b)(-1)^k$$

$$\Gamma_k = (a-b)\,(-1)^k\,\frac{a^kb - a^{k-1}}{a^2-1} \tag{78}$$

$$\Gamma_0 = \left(\sum_{i=1}^{\infty} a^{i-1}(a-b)^2a^{i-1}(-1)^{2i-2}\right) + 1 = \frac{2ba-b^2-1}{a^2-1} \tag{79}$$

Then,

$U(z) = \sum_{k=1}^{\infty} (a-b)\,(-1)^k\,\frac{a^kb-a^{k-1}}{a^2-1}z^{-k} + \frac{2ba-b^2-1}{a^2-1} = \frac{ba^2-b+2baz-b^2z-z}{(a^2-1)(a+z)}$

The corresponding tensor is:

$$g_{11} = \\ = \frac{3b^2a^6-4b^3a^5-4a^5b-7b^2a^4+24a^3b+24b^3a^3-49b^2a^2-9a^2-9a^2b^4+20b^3a+20ba-1-b^4-7b^2}{(a-1)^5(a+1)^5}$$

$$g_{12} = -\frac{3a^5b-2a^4-4b^2a^4-6a^3b+7a^2+17b^2a^2-15ba-6b^3a+1+5b^2}{(a-1)^4(a+1)^4}$$

$$g_{22} = \frac{3a^4-4a^3b-6a^2+12ba-1-4b^2}{(a-1)^3(a+1)^3} \tag{80}$$

6 Conclusions

The differentiable manifold of systems of a common McMillan degree is a natural set for the state space representation, because this degree is the number of components of the state in the minimal representation, thus appearing explicitly in the models. It gives rise to the so-called overlapping parametrizations, which exhibit much more flexibility than the canonical ones, which is important for the numerical process of identification, given the possibility of changing of model even on-line, whenever a malconditioning is detected.

For the ARMA representation, the establishing of overlapping

parametrizations based upon classes of models representing systems with a common McMillan degree results in a clumsy class of mathematical models [4], because this degree is not natural for that representation. Unfortunately, the natural integer - p, the minimum degree of the AR and MA polynomials - doesn't give rise, as shown, to a class of models whose image (the systems of a common p) is a differentiable manifold, thus not allowing the definition of a natural ARMA overlapping parametrization.

A remedial solution could be to work with a rougher class of models (the ARMA(p,p)-irreducible ones), which would not be strictly identifiable for systems in which not all Kronecker indices are equal; since the set of systems in which all them are equal is generic, maybe this was not such a big handicap...

The complexity of the Riemannian metric tensor formulas grow exponentially with the increase of systems dimensions. There are two ways of attenuating this problem:

1) Computer languages for algebraic symbolic processing like MAPLE and MATHEMATICA^{TM} are sufficiently flexible to integrate numerical algorithms with symbolic ones, thus dispensing the need of manual transcription of formulas.

2) Since the only systems treated in the theory here presented are the stable ones, the parameters that appear in the denominator of the transfer function are bounded by restriction such as being between -1 and 1. In this case, certain terms of higher order in the expressions obtained for the metric tensors can be despised without considerable loss of precision, thus reducing the complexity of the formulas.

The availability of powerful computer packages for algebraic computation turns attractive the use of analytical formulas involving complex integrals and partial derivatives of polynomial matrices for the study of geometrical properties of spaces of linear dynamical systems.

Given the close relations between the Riemannian metric tensor, the Fisher information matrix, the covariance matrix of the parameters estimators and the Hessian matrix of some common objective functions used in parametric identification, the author believes that the results displayed in this article hold some relevance for the classical problem of linear dynamical systems identification.

References

[1] Aoki, M. (1987). **State space modeling of time series**. Springer Verlag.

[2] Aplevich, J.D. **Singular Pencil Models in Systems Design and Control**, Internal Report N2L 3G1, Elet. Eng. Dpt., University of. Waterloo, 1981.

[3] Box, G.E.P. and Jenkins,G.M. **Time Series Analysis, Forecasting and Control**. Holden-Day. San Francisco. 1970.

[4] Correa, G. O. and Glover, K. **Pseudo-canonical forms, identifiable parametrizations and simple estimation for linear multivariable systems: input-output models.** Automatica, vol. 20, n.4, pp. 429-442. 1984.

[5] Chou, T. C., **Geometry of Linear Systems and Identification**. Ph.D. thesis. Trinity College, Cambridge,1994.

[6] Clark, J.M.C.**The consistent selection of local coordinates in linear systems identification**, JAAC Purdue University, Lafayette, Indiana, 1976, 576-580.

[7] Denham, M. J. (1974). **Canonical forms for the identification of multivariable linear systems**. IEEE Trans. Autom. Control. Vol AC-19, n.6, pp. 646-656.

[8] Guidorzi, R.P. **Invariants and canonical forms for systems structural and parametric identification**. Automatica, vol.17, n.1,pp117-133.

[9] Goodwin, G. C. e Payne, R.L. (1977). **Dynamic system identification**. Academic Press.

[10] Gevers, M. and Wertz,V. **Uniquely identifiable state-space and ARMA parametrizations for multivariable systems**, Automatica, vol. 20, n. 3, 1984, 333-347.

[11] Hannan, E.J. and Deistler, M. **The statistical theory of linear systems**. John Wiley and Sons. 1988.

[12] Hanzon, B. **Identifiability, recursive identification and spaces of linear dynamical systems**, Ph.D. Thesis, Department of Econometrics, Erasmus University, Rotterdam, 1986.

[13] Kailath, T. **Linear Systems**. Prentice Hall.1980.

[14] Oppenheim, A.V. and Shafer, R.W. **Digital signal processing**, Prentice-Hall, 1976, 66.

[15] Peeters, R. **System identification based on Riemannian geometry: theory and algorithms**, Ph.D. Thesis, Free University of Amsterdam, 1994, and Research Report nr. 64, Tinbergen Institute Research Series, Tinbergen Institute, Rotterdam.

[16] Tiao, G.C. and Tsay, R.S. **Multiple time series modeling and extended sample cross correlations**. Technical report n. 690. Statistics Dept. Univesity of Wisconsin. 1982.